Faulkner SC, Verity R, Roberts D, Roy S, Robertson PA, Stevenson MD, Le Comber S.

Using geographic profiling to compare the value of sightings vs trap data in a biological invasion.

***Diversity and Distributions* 2016**

Using geographic profiling to compare the value of sightings vs trap data in a biological invasion

Sally C. Faulkner[1], Robert Verity[2], David Roberts[3], Sugoto S. Roy[4], Peter A. Robertson[5,6], Mark D. Stevenson[§,1] & Steven C. Le Comber[§,1].

[1]School of Biological and Chemical Sciences, Queen Mary University of London, London. E1 4NS, UK.
[2]Medical Research Council Centre for Outbreak Analysis and Modelling, Imperial College London, London. W2 1PG, UK.
[3]Durrell Institute of Conservation and Ecology, School of Anthropology and Conservation, University of Kent, Marlowe Building, Canterbury, Kent. CT2 7NR, UK.
[4]International Union for the Conservation of Nature, 28 re Mauverney, CH-1196 Gland, Switzerland.
[5]School of Biology, Newcastle University. NE1 7RU, UK.
[6]Animal and Plant Health Agency, Sand Hutton, York. YO41 1LZ, UK.

[§]Corresponding authors: s.c.lecomber@qmul.ac.uk, mds_mds@live.com

Abstract

Aim - The development of conservation plans, including those dealing with invasive species, is underpinned by the need to obtain reliable and accurate data. However, in many cases responding rapidly is equally critical.

Location - The data were obtained from the Hebridean Mink Project, which was set up with the objective of removing mink from North Uist, Benbecula and South Uist.

Methods - Here, we introduce an extension of the Dirichlet Process Mixture (DPM) model of geographic profiling that can be used to estimate source locations of invasions directly from spatial point pattern data without the need to specify dispersal parameters. We use this model to analyse a biological invasion of American mink (*Neovison vison*) in the Hebrides.

Results - Our results suggest that sightings data – which are relatively easy and quick to acquire – can be used to capture much of the information about sources of invasive species that is obtained from the harder to acquire and more intensive trap data.

Main conclusion - These results have important implications for the development of conservation plans and, in this case, in the early stages of biological invasions, when interventions are most likely to be successful.

Key words: Bayesian models, citizen science, conservation management, geographic profiling, invasive species, mink

Introduction

In many areas of biology – for example wildlife conservation, or invasion biology– the development of management strategies is hampered by the lack of reliable and accurate data (Tyre et al. 2003; Faulkner et al. 2015). However, in many cases responding rapidly is critical – perhaps more so than waiting to obtain high quality data – since delaying action until further data is obtained can be catastrophic and can hinder opportunities to prevent species extinctions (Grantham et al. 2009, Martin et al. 2012). Thus, extracting the maximum amount of information from existing data is crucial, even when the data sources are less than ideal – for example, ad hoc sightings rather than systematic survey data. Traditional survey techniques can often be limited by time and financial constraints, along with logistical challenges that are presented by particular habitats (Faulkner et al. 2015). In addition detection of a species is not always possible until the species has reached a particular threshold, and this detection threshold will depend not only on how well established the species is but also on the monitoring method used (Myers et al. 2000; Harvey et al. 2009; Dejean et al. 2012).

One instance in which rapid action may be especially important is invasive species management. Invasive species are a major factor affecting global diversity and ecological integrity, with costs totaling millions of dollars per episode (Vitousek et al. 1996; Wilcove et al. 1998; Gurevitch & Padilla 2004; Didham et al. 2005; Blackburn et al. 2011; Pyšek et al. 2012; Robertson et al. 2015). The cost of the response to invasive species incursion is best predicted by the area over which the species has spread (Martins et al. 2006; Pluess et al. 2012; Robertson et al. 2015). Estimating this area at an early stage is therefore important to reduce costs. Damage due to invasive species is particularly prevalent within island ecosystems, where species have evolved in the absence of strong ecological competition (Nogales et al. 2006). In addition island environments are often characterised by high levels of endemism, or distinctive ecological processes that have a high potential to be disrupted by invasive mammals (Stone et al. 1994; Nogales et al. 2006). However, it is important to note that not all invasive species are deleterious, and that there are some positive interactions that can occur which drive the diversity-invasibility relationship. These interactions are particularly important when dealing with issues such as climate change, which makes it even harder to identify invaders (Altieri et al. 2010).

The American mink (*Neovison vison*) is a widely distributed invasive carnivore now found within 28 European countries, and is one of five non-native mammals that have established feral populations in Europe (Bonesi and Palazon 2007; Roy et al. 2015). It can reach high population densities and is known to have a major negative impact on ground nesting birds and other native fauna (Roy et al. 2011). Management efforts have proved challenging with most strategies concentrating on long term control operations rather than eradications (Bonesi and Palazon 2007; Roy et al. 2015). Feral mink populations are also establishing in vulnerable island populations such as the Outer Hebrides in Scotland. Here they established following the release or escape from fur farms on the Isle of Lewis in the 1950s, after which they then spread south on to Harris and the Uists (Roy et al. 2015).

Early detection, rapid assessment and rapid response are crucial for preventing invasive species establishment and spread (National Invasive Species Council 2008). Even when there is only limited knowledge of an invasive species' population ecology, it can often be eradicated early on in the invasion. For this reason, financial investments are most beneficial during the early stages of the search for an invader (Holden et al. 2016). However, much of the invasive species literature is focused on costly population biology studies and determining the best control strategies under complete knowledge of invader dispersal and demography (Simberloff 2003; Epanchin-Niell & Hastings 2010; Holden et al. 2016). Given limited conservation budgets cost efficacy is important for achieving the best outcomes (Naidoo et al. 2006) and therefore studies on dispersal processes may only be necessary when the window of opportunity for early detection and eradication is lost (Simberloff 2003). Consequently, methods that can be used quickly on limited data sets (for example citizen science data) can be important, potentially decreasing the financial strain of invasive species management

As in many similar examples, the mink invasion highlights the time and expense required to acquire high-quality trapping data. One tempting alternative is to use sightings data produced by members of the public as part of 'citizen science', since this is likely to be quicker, easier and less costly to collect. Citizen science is the process by which the public are involved in science as researchers (Keough and Blahna 2006; Conrad and Hilchey 2011) – although of course 'citizen science' in the form of the

naturalist was established long before academic research became a form of employment. Advocates of citizen science in its modern sense claim a number of benefits (for example, the 'democratisation' of science, or the acquisition of new skills to non-professional volunteers) although fewer provide empirical evidence (Legg and Nagy 2006; Conrad and Hilchey 2011). More recently, citizen-based observation data have been viewed as an important tool for investigating spatial patterns of a wide range of taxa and natural processes (Sullivan et al. 2009; Snäll et al. 2011). However, a number of concerns have been raised about the quality of data collected via citizen science (Tyre et al. 2003). In this study, we use geographic profiling to compare the results obtained from analysis of (a) professional trap data, and (b) sightings data derived from citizen science.

Geographic profiling (GP) is a statistical technique originally developed in criminology to prioritise large lists of suspects in cases of serial crime such as murder, rape and arson by using the spatial locations associated with linked crimes to calculate the probability of offender residence for each point in the study area (Rossmo 2000); suspects are then selected for further investigation according to their height on the resultant probability surface. More recently this method has been applied to biological data, including animal foraging behaviour (finding animal nests or roosts using the locations of foraging sites as input) (Le Comber et al. 2006; Martin et al. 2009; Raine et al. 2009; Le Comber & Stevenson 2012), epidemiology (identifying disease sources from the addresses of infected individuals) (Buscema et al. 2009; Le Comber et al. 2011), conservation (locating the sleeping sites of rare, nocturnal tarsiers using locations at which calls were detected) (Faulkner 2015) and invasive species biology (using current locations to identify source populations) (Stevenson et al. 2012). Using historical data from the Biological Records Centre, Stevenson et al. (2012) were able to show that GP performed significantly better both than analyses using measures of spatial central tendency, and a more complex single-parameter kernel density model. At the same time the model has been adapted to fit a Bayesian framework in the form of a Dirichlet Process Mixture (DPM) model which is designed to deal with situations with multiple sources when even the number of sources is unknown (Verity et al. 2014). The model can also be used to estimate dispersal parameters (as in Le Comber et al. 2006).

In this study, we ask whether GP can be applied to the type of data-poor situations that are likely to characterise the early stages of biological invasions when data quality may be low; for example, when sample sizes are small, when data are based on public sightings data rather than systematic trap data, when locations are not precisely recorded, when identification is uncertain or when there is limited knowledge about the way in which the species in question might disperse across novel habitat. To deal with this last point, we develop a novel mathematical extension to the DPM model that requires minimal input from the user about the likely dispersal of the invading species, testing this on both simulated and real-world data sets. Specifically, we ask: (i) does the model extension introduced here lead to correct posterior inference of dispersal parameters when tested with simulated data? (ii) can this novel implementation of the DPM model be used to locate known den sites associated with the invasion of American mink in the Hebrides, using either trap data or sightings data as 'crime sites'?; (iii) how does the geoprofile derived from public sightings data compare to that derived from professional trap data?

Methods

Note that although the terminology used in GP – for example crime sites, offender residence – is not always appropriate for use with ecological and biological data, for consistency this paper will refer to mink sightings/captures as crime sites and mink dens/breeding grounds as source locations. No negative connotations should be inferred from this.

The DPM model

The model used here is an extension of the DPM model described by Verity et al. (2014). The original method works by breaking down the difficult problem of estimating multiple unknown source locations into two conceptually simpler sub-problems. The first step is to estimate which "crimes" originated from which sources based on the locations of these sources – which are assumed known at this stage. Observations are more likely to be allocated to sources that are close by, leading to a clustering of observations into a finite number of groups, where members of the same group are assumed to have originated from the same source. The second step is to re-estimate source locations based on the clustering, which is now assumed known. The

final model implementation uses Bayesian Markov chain Monte Carlo (MCMC) to alternate between these steps many thousands of times until the algorithm converges on the posterior distribution of interest, with convergence being measured via the Gelman-Rubin diagnostic statistic (Gelman et al. 2004). These posterior draws can be used to generate a probability surface of the unknown sources (i.e. a geoprofile) together with an estimate of the number of sources from which the observed data originated. Further details of the method can be found in Verity et al. (2014).

Fitting σ from point pattern data

In the original implementation of the model described in Verity et al. (2014) (and used in Faulkner et al. (2015)) the dispersal distance was defined by the fixed parameter σ, which had to be set by the user. σ represents the standard deviation of the bivariate normal distribution describing dispersal from a central source. Previously, this was in units of degrees of latitude/longitude; in the new version of the model presented here, we use km to avoid differences in distance between 1 degree latitude and 1 degree longitude. So, a σ value of 1 km will imply that 39% of dispersal events occur within 1 km of the source, 87% within 2 km and 99% within 3 km. In many cases, existing knowledge of the species in question will suffice to provide a reliable estimate of σ. However, in the case of invasions this may be more challenging, due to the uncertainty associated with a non-native species moving into a new and atypical habitat. In such situations the ability to estimate σ from point pattern data alone will clearly be advantageous, and it is this extension to the DPM model that we introduce here.

The underlying mathematics of this extension are described in Appendix 1, but in brief the method uses a Gibbs sampler to estimate the locations of the sources and the value of σ simultaneously. We assume an inverse-gamma prior on σ^2, which has the advantage of being flexible enough to capture broad uncertainty in σ, while also being mathematically convenient due to being conjugate to the likelihood. The user is required to specify two parameters that control the mean and shape of this distribution. Generally, our approach is to use values such that the prior is extremely diffuse and heavy-tailed, and thus is relatively uninformative of σ.

Model implementation

The DPM model described here was implemented in R (R Core Team 2014) using a

modification of the package Rgeoprofile as described by Verity et al. (2014); this package is available at https://github.com/bobverity/Rgeoprofile.

Testing the model

Model output can be assessed using the hit score: the proportion of the area covering the crimes (here, as is common in criminology, the positions of the 'crimes', plus a 'guard rail' of 5% surrounding this). The hit score is calculated by dividing the ranked score of each grid square within the total search area by the total number of grid squares. The smaller the hit score, the more efficient the search strategy described by the model; a hit score of 0.5 (50%) is what would be expected from a non- prioritized (i.e. random or systematic) search (Rossmo 2000).

Comparing geoprofiles

For both analyses, the study area was defined as the rectangular area encompassing all trap and sightings data, plus a 'guard rail' of 5%. Since correlations between geoprofiles produced from the two data sets would be artificially inflated by the low scores assigned to the sea in both cases (where there are neither sightings nor trap data), even if the two approaches produced quite different results for the areas of interest (the land), we took the conservative approach of calculating Pearson's product moment correlation coefficients after restricting the analysis to the top 50% of the geoprofile in each case (largely corresponding to land, and shown in Figure 2). This ensured that distributions were compared in terms of high priority regions.

Data

The data were obtained from the Hebridean Mink Project, which was set up with the objective of removing mink from North Uist, Benbecula and South Uist, whilst also trying to reduce the density from South Harris. Trapping started in 2001, and data were collected until 2005. The area used in the study included a mixture of saltwater and freshwater habitats – 1,116km of coastline, 2,416km of loch shore and 189km of freshwater rivers and streams. Full details of the project can be found in Roy et al. (2011, 2015).

Sightings and trap data

The data consisted of 125 sightings identified by the public (in many cases, crofters) over four years (2002-2005) and 409 total captures over five years (2001-2005, 141,498 trap nights). These data covered both Uist and Harris. Locations for 15 known dens (sources) were used for assessing the performance of the model.

Simulations

To test the accuracy of the model, 1000 simulated data sets were created, with 1-5 sources (drawn from a uniform distribution between -4 and -3º longitude, and 56 and 57º latitude), with 5-10 crimes per source, again drawn from a uniform distribution. The actual value of σ was set independently for each simulation, drawing from a uniform distribution ranging from 1-10 km. The DPM model was run on each data set, with the prior expectation of σ chosen independently but in the same way, and using the same range of values; these values, combined with the area from which the sources were selected, meant that crimes from different sources often overlapped, providing a more challenging (and more realistic) test for the model. The shape parameter of the inverse-gamma prior on σ^2 was set to 2, corresponding to a weakly informative distribution. In each simulation the prior expectation of σ, the posterior mean value of σ, and the true value of σ were recorded.

Results

Simulations

Across 1000 replicates, the fitted value of σ was most strongly affected by the actual value, although the prior expectation of σ, the number of sources and the number of crimes also had significant effects; the fitted coefficient for the relationship between actual and fitted values was 0.92, with an intercept of -1.35 (ANOVA: prior expectation $F_{1,995} = 114.5$, $p < 0.0001$; sources $F_{1,995} = 15.8$, $p < 0.0001$, crimes $F_{1,995} = 123.2$, $p < 0.0001$; actual value $F_{1,995} = 2799.1$, $p < 0.0001$) (Fig. 1). Model error, expressed as the difference in km between the actual σ value and the fitted value, increased with number of sources, with a slight tendency to underestimate σ when there was a single source and to overestimate with multiple sources ($F_{4,995} = 21.5$, $p < 0.0001$) (Fig. 1c).

Profiling the trap and sighting data

The geoprofiles for the trap and sightings data are shown in Fig 2 and the hit scores for known den locations in Table 1. Using the trap data, the mean hit score for the 15 known den sites was 0.06 (sd 0.058) and the median was 0.04, with the model finding all of the known dens by searching just over 20% of the study area, 13 by searching 10% and four less than 1%. Using the sightings data the mean hit score was 0.14 (sd 0.124; median 0.10), with the model finding eight of the dens by searching less than 10% of the total search area. Although the hit scores for the trap data were lower (paired t-test: $t = 2.9$, $df = 14$, $p = 0.01$), overall the geoprofiles derived from the trap data and the sightings were similar (Pearson's product moment correlation: $r = 0.42$, $p < 0.0001$), despite the discrepancy in the size of the two data sets (409 data points for the trap data versus 125 sightings).

Discussion

Our study shows for the first time that citizen science data – in the form of sightings – and the DPM model could together be used to target interventions and estimate the area required for control in the early stages of an invasion, when control efforts are most likely to be effective. In addition, our extension of the DPM model enhances the model's utility in data poor situations that might characterise the movement of an invasive species into a novel habitat.

GP's origins lie in criminology, where it was developed as a tool to prioritise suspect lists in cases of serial crime, since either time or resources often preclude detailed investigation of the large lists of suspects that typify such investigations (for example, 268,000 suspects and 5.4 million vehicle registrations in the case of the Yorkshire Ripper (Doney 1990)). Similar problems are also encountered in conservation, since time and resources are likely to be similarly limited (Stevenson et al. 2012). Added complications that apply in both criminology and conservation are (i) sample sizes may be low (models depending on large sample sizes are obviously useless when data points are murders), and (ii) that data may be incomplete (i.e. there may be unknown or unrecorded crimes, or populations of an invasive species). GP is generally robust to both of these problems, and can be successful on as few as five locations (Rossmo 2000). However, in conservation if not criminology, there may be the option of using larger data sets even if the data are of lower quality – for example, data derived from

citizen science rather than rigorous survey or trap data. Here, we show that in the case of the American mink in the Hebrides GP works very well at identifying known den sites using trapping data or when using sightings and – crucially – that the results of each are similar. The similarity of the results is particularly impressive since the sightings data set was smaller than the trap data, in addition to being (presumably) of lower quality.

Our study also introduces an extension to the DPM model of GP originally described in Verity et al. (2014). This original paper placed GP within a Bayesian framework, while also allowing for multiple sources, even when the number of sources is unknown (previous Bayesian models had explicitly assumed a single source) (O'Leary 2009, 2010a, 2010b, 2012). However, the model did require the user to specify a fixed parameter, σ, describing the standard deviation (in that version of the model, in decimal degrees of latitude/longitude) of the bivariate normal distribution describing dispersal from the source. The version of the DPM model introduced here removes this limitation, and requires only an expected value of σ (now in km) and a prior shape parameter.

There are a number of other ways in which the basic model used here could be extended. Rather than a normal dispersal distribution, other types of distributions (for example, the Cauchy distribution or Student's t distribution) could be used, and in particular distributions that are biased in particular directions, to account for the fact that dispersal may be easier in some directions than others, perhaps because of prevailing winds etc. In addition, more informed priors could be used – in the data set described here, for example, setting a zero prior on the sea, eliminating the need for the (cruder) *post hoc* correction used here.

The results of the simulations show good correspondence between the actual σ value and the fitted value (Fig 1b). The prior expectation of σ had a significant effect on the model (Fig 1a), although eventually we would expect the prior to be overwhelmed by the data as the number of 'crimes' increased (in simulations there were as few as five crimes); in any case, this effect was minor compared to the effect of the actual σ value, with an F statistic of 114.5 compared to 2799.1. The model had a slight tendency to overestimate σ as the number of sources increased, almost certainly due to overlapping

distributions of crimes from closely adjacent sources leading the model to select a single, larger cluster (as Fig 1c shows, simulations with single sources typically have lower fitted values of σ than simulations with multiple sources). Again, this effect was small compared to the effect of the true value, with an F statistic of 15.8.

Together, the use of sightings data and the removal of the requirement to specify σ in the DPM model suggest that this approach could potentially be of use in other biological invasions, particularly in the early stages when interventions are likely to be most successful but when data quality may be low and the population poorly characterised. With the cost of invasive species running from millions to billions of dollars per occurrence (Vitousek & D'Antonio 1996; Wilcove et al. 1998; Gurevitch & Padilla 2004; Didham et al. 2005; Blackburn et al. 2011; Pyšek et al. 2012), it is clear that tools helping to inform management strategy in real time are highly advantageous, and we suggest that this implementation of the DPM model provides one such tool. In particular, targeting control efforts in this way will reduce the area required for control, making interventions more efficient and more cost effective.

Acknowledgements

The data collection was funded by the EU-LIFE programme, Scottish Natural Heritage, The Scottish Executive, The Royal Society for the Protection of Birds, Western Isles Council and Western Isles Enterprise. The authors would like to express their gratitude to the landowners for allowing access to their land, to foremen and the trappers for their hard work and to those who reported mink sightings.

References

Altieri, A. H., van Wesenbeeck, B. K., Bertness, M. D. and Silliman, B. R. (2010) Facilitation cascade drives positive relationship between native biodiversity and invasion success. *Ecology*. **91**, 1269–1275.

Blackburn, T.M, Pyšek, P., Bacher, S., Carlton, J.T., Duncan, R.P., Jarošík, V., Wilson, J.R.U. & Richardson, D.M. (2011) A proposed unified framework for biological invasions. *Trends in Ecology & Evolution*. **26**, 333-339.

Bonesi, L. & Palazon, S. (2007). The American mink in Europe: Status, impacts, and control. *Biological Conservation*. **134**, 470-483.

Buscema, M., Grossi, E., Breda, M. & Jefferson, T. (2009). Outbreaks source: A new mathematical approach to identify their possible location. *Physica A: Statistical Mechanics and its Applications*. **388**, 4736–4762.

Conrad, C. C. & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment*. **176**, 273-91.

Dejean, T., Valentini, A., Miquel, C., Taberlot, P., Bellemain, E. & Miaud C. (2011) Improved detection of an alien invasive species through environmental DNA barcoding: the example of the American bullfrog *Lithobates catesbeianus. Journal of Applied Ecology*. **49**, 953-959.

Didham, R.K., Tylianakis, J.M., Hutchison, M.A., Ewers, R.M. & Gemmell, N.J. (2005) Are invasive species the drivers of ecological change?. *Trends in Ecology & Evolution*. **20**, 470-474.

Doney, R.H. (1990) The aftermath of the Yorkshire Ripper: The response of the United Kingdom police service. In: S.A. Egger (ed.), Serial Murder: An Elusive Phenomenon. New York, NY: Praeger.

Epanchin-Niell, R. S. and Hastings, A. (2010) Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecology Letters*, **13**, 528–541.

Faulkner, S.C., Stevenson, M.D., Verity, R., Mustari, A.H., Semple, S., Tosh, D.G. & Le Comber, S.C. (2015) Using geographic profiling to locate elusive nocturnal animals: a case study with spectral tarsiers. *Journal of Zoology*. **295**, 261-268.

Gelman, A., Carlin, J.B., Stern, H.S., & Rubin, D.B. (2004) Bayesian Data Analysis, 2nd edition, Chapman & Hall/CRC, Boca Raton, Florida.

Grantham, H.S., Wilson, K.A., Moilanen, A., Rebelo, T. & Possingham, H.P. (2009) Delaying conservation actions for improved knowledge: how long should we wait?. – *Ecology Letters*. **12**, 293-301.

Gurevitch, J. & Padilla, D.K. (2004) Are invasive species a major cause of extinctions?. *Trends in Ecology & Evolution*. **19**, 470-474.

Harvey, C.T., Qureshi, S.A. & Maclsaac, H.J. (2009) Detection of a colonizing, aquatic, non-indigineous species. *Diversity and Distribution*. **15**, 429-437.

Holden, M.H., Nyrop, J.P. and Ellner, S.P. (2016) The economic benefit of time-varying surveillance effort for invasive species management. *Journal of Applied Ecology*.**53**, 712-721

Keough, H. & Blahna, D. (2006) Achieving integrative, collaborative ecosystem management. *Conservation Biology*. **20**, 1373-1382.

Le Comber, S.C. & Stevenson, M.D. (2012) From Jack the Ripper to epidemiology and ecology. *Trends in Ecology & Evolution*. **27**, 307.

Le Comber, S.C., Nicholls, B., Rossmo, D.K. & Racey, P,A. (2006) Geographic profiling and animal foraging. *Journal of Theoretical Biology*. **240**, 223-240.

Le Comber, S.C., Rossmo, D.K., Hassan, A.N., Fuller, D.O. & Beier, J.C. (2011) Geographic profiling as a novel spatial tool for targeting infectious disease control. *International Journal of Health Geographics*. **10**, 35.

Legg, C. & Nagy, L. (2006) Why most conservation monitoring is but need not be, a waste of time. *Journal of Environmental Management*. **78**, 194-199.

Martin, T.G., Nally, S., Burbidge, A.A., Arnall, S., Garnett, S.T., Hayward, M.W., Lumsden, L.F., Menkhorst, P., McDonald-Madden, E., Possingham, H.P. (2012) Acting fast helps avoid extinction. *Conservation Letters.* **5,** 274-280.

Martin, R.A., Rossmo, D.K. & Hammerschlag, N. (2009) Hunting patterns and geographic profiling of white shark predation. *Journal of Zoology*. **279**, 111-118.

Martins, T.L.F., Brooke, M.deL., Hilton, G.M., Farnsworth, S., Gould, J. & Pain, D.J. (2006) Costing eradications of alien mammals from islands. *Animal Conservation*. **9**. 439-444 .

Myers, J.H., Simberloff, D., Kuris, A.M. & Carey, J.R. (2000) Eradication revisited: dealing with exotic species. *Trends in Ecology & Evolution*. **15**, 316-320.

Naidoo, R., Balmford, A., Ferraro, P.J., Polasky, S., Ricketts, T.H., Rouget, M. (2006) Integrating economic costs into conservation planning. *Trends in Ecology and Evolution.* **21**, 681-687.

National Invasive Species Council (2008) 2008–2012 National Invasive Spe- cies Management Plan. Department of the Interior, Washington, DC.

Nogales, M., Rodriguez-Luengo, J.L. & Marrero, P. (2006) Ecological effects and distribution of invasive non-native mammals on the Canary Islands. *Mammal Review*. **36**, 49-65.

O'Leary, M. (2009) The mathematics of geographic profiling. J*ournal of Investigative Psychology and Offender Profiling*. **6**, 253–265.

O'Leary, M. (2010a) Implementing a Bayesian approach to criminal geographic profiling. First international conference on computing for geospatial research and application, June 21–23, Washington, DC.

O'Leary, M. (2010b) Multimodel inference and geographic profiling. *Crime Mapping*. **2**.

O'Leary, M. (2012) New mathematical approach to geographic profiling. National Institute of Justice, Washington, D.C.

Pluess, T., Cannon, R., Jarošík, V., Pergl, J. & Bacher S. (2012) When are eradication campaigns successful? A test of common assumptions. *Biological Invasions*. **14**, 1365-1378.

Pyšek, P., Jarošík, V., Hulme, P.E., Pergl, J., Hejda, M., Schaffner, U. &Vilà M. (2012) A global assessment of invasive plant impacts on resident species, communities and ecosystems: the interaction of impact measures, invading species' traits and environment. *Global Change Biology*. **18**, 1725-1737.

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing , <www. R-project.org>.

Raine, N.E., Rossmo, D.K. & Le Comber, S.C. (2009) Geographic profiling applied to testing models of bumble-bee foraging. *Journal of the Royal Society Interface*. **6**, 307–319.

Robertson, P.A., Adriaens, T., Lambin, X., Mill, A., Roy, S., Shuttleworth, C.M. & Sutton-Croft, M. (2015) The large-scale removal of mammalian invasive alien species in Northern Europe. *Pest Management Science* DOI: 10.1002/ps.4224.

Rossmo, D.K. (2000) Geographic profiling. – CRC Press.

Roy, S.S. (2011) Strategies to improve landscape scale management of mink populations in the west coast of Scotland: lessons learned from the Uists 2001-2006. Island Invasives: Eradication and Management. IUCN, Gland, Switzerland.

Roy, S.S., Chauvenet, A.L.M. & Robertson, P.A. (2015) Removal of Americal mink (*Neovison vison*) from the Uists, Outer Hebrides, Scotland. *Biological Invasions*. **17**, 2811-2820.

Simberloff, D. (2003) How Much Information on Population Biology Is Needed to Manage Introduced Species?. *Conservation Biology,* **17**, 83–92.

Snäll, T., Kindvall, O., Nilsson, J. & Pärt, T. (2011) Evaluating citizen-based presence data for bird monitoring. *Biological Conservation*. **144**, 804-810.

Stevenson, M.D., Rossmo, D.K., Knell, R.J. & Le Comber, S.C. (2012) Geographic profiling as a novel spatial tool for targeting the control of invasive species. *Ecography*. **35**, 704-715.

Stone, P.A., Snell, H.L. & Snell, H.M. (1994) Behavioural diversity as biological diversity: Introduced cats and lava lizard wariness. *Conservation Biology*. **8**, 569-573.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D. & Kelling, S. (2009) eBrid: A citizen-based bird observation netwrok in the biological sciences. *Biological Conservation*. **142**, 2282-2292.

Tyre, A.J., Tenhumberg, B., Field, S.A., Niejalke, D., Parris, K. & Possingham, H.P. (2003) Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*. **13**, 1790-1801.

Verity, R., Stevenson, M.D., Rossmo, D.K., Nichols, R.A. & Le Comber, S.C. (2014) Spatial targeting of infectious disease control: indentifying multiple, unknown sources. *Methods in Ecology and Evolution*. **5**, 647-655.

Vitousek, P.M., D'Antonio, C.M., Loope, L.L. & Westbrooks, R. (1996) Biological invasions as global environmental change. *American Scientist*. **84**, 468-478.

Wilcove, D.S., Rothstein, D., Dubow, J., Philips. A. & Losos E. (1998) Quantifying threats to imperiled species in the United States. *Bioscience*. **48**, 607-615.

Biosketch

The researchers involved in this analyses work in or have collaborated with the Le comber lab (https://evolve.sbcs.qmul.ac.uk/lecomber/). This lab covers a wide range of subjects within evolutionary biology, with an emphasis on developing geographic profiling techniques and applying them to biological data.

Author contributions: DR, SSR, SCF and SLC conceived the idea. SCF and SLC proceeded with the data analysis and led the writing. SSR and PAR provided the data. RV, MDS and SLC led the development of the model. All authors contributed to the final written production.

Appendix 1

Here we build on the basic DPM model that is described in detail in the appendix of Verity et al. (2014) to allow for variable dispersal distance. Following the same notation, and assuming that crime sites $\boldsymbol{x}_i$ (for $i \in 1, \ldots, n$) are drawn from a bivariate normal distribution around their corresponding source locations $\boldsymbol{\mu}_{c_i}$, we obtain the following likelihood:

$$f(\boldsymbol{x}\,|\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^n \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}_{c_i}\right)^T\left(\boldsymbol{x}_i - \boldsymbol{\mu}_{c_i}\right)\right\} \tag{1}$$

(where $\boldsymbol{x}_i$ and $\boldsymbol{\mu}_{c_i}$ are column vectors, each with two rows). In order to infer σ^2 we must first choose a suitable prior for this parameter. We follow the convention of using an inverse-gamma prior on σ^2:

$$f(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\delta+1} \exp\left\{\frac{-\beta}{\sigma^2}\right\}. \tag{2}$$

This prior has the significant advantage of being conjugate to the likelihood; if the prior distribution on σ^2 is written Inv-Gamma(δ, β) then the conditional posterior is given by Inv-Gamma$\left(\delta + n,\ \beta + \frac{1}{2}\sum_{i=1}^{n}\left(\boldsymbol{x}_i - \boldsymbol{\mu}_{c_i}\right)^T\left(\boldsymbol{x}_i - \boldsymbol{\mu}_{c_i}\right)\right)$. By drawing a new value of σ^2 from this conditional posterior distribution at each iteration of the MCMC – conditional on known source locations ($\boldsymbol{\mu}$) and cluster allocation ($\boldsymbol{c}$) at that point in the algorithm – we obtain a series of draws from the complete posterior distribution of σ^2. The posterior distribution of source locations (i.e. the geoprofile) can then be considered to be integrated over the prior on σ^2. We note that in the original Verity et al. (2014) application $\boldsymbol{\mu}$ was integrated out analytically, meaning only the group allocation needed to be updated as part of the MCMC sampler. Here we can still make use of this gain in efficiency by integrating over $\boldsymbol{\mu}$ when updating the group allocation, however, $\boldsymbol{\mu}$ values must be realized immediately prior to updating σ^2, as the posterior on σ^2 is conditional on these values.

Finally, when defining prior parameters we choose to work in the space of σ rather than σ^2. The parameters set by the user are the expectation of σ, denoted ε, and the shape parameter δ. The distribution of σ can then be written

$$f(\sigma) \propto \left(\frac{1}{\sigma}\right)^{2\delta+1} \exp\left\{-\left(\frac{\varepsilon\,\Gamma(\delta)}{\sigma\,\Gamma\left(\delta - \frac{1}{2}\right)}\right)^2\right\}.$$

The advantage of this setup is that ε has an intuitive interpretation as the expected

dispersal distance. The prior variance of σ is given by

$$\mathrm{Var}(\sigma) = \varepsilon^2 \left(\frac{\Gamma(\delta - 1)\Gamma(\delta)}{\Gamma(\delta - \frac{1}{2})^2} - 1 \right).$$

TABLES

Table 1. Latitude and longitude coordinates of known dens, with hit scores of search area (km^2) for the two analyses.

Latitude	Longitude	Hit score		Area Searched (km^2)	
		Trap data	Sightings data	Trap data	Sightings data
-6.8645701	57.794056	<0.001	0.013	0.2	62.0
-7.1439991	57.652154	0.003	0.058	15.1	284.6
-7.0150114	57.758104	0.003	0.175	15.3	863.9
-7.1384491	57.657167	0.004	0.071	18.3	350.4
-6.8578971	57.789041	0.013	0.033	65.6	164.6
-7.1339153	57.663083	0.020	0.095	100.5	467.4
-7.1550095	57.665138	0.040	0.106	197.0	523.7
-7.3170788	57.177832	0.043	0.064	211.6	316.2
-7.1289061	57.663277	0.050	0.096	245.2	473.4
-7.0513014	57.655395	0.055	0.138	269.4	681.1
-7.289443	57.701967	0.063	0.262	310.2	1294.8
-7.3499298	57.271959	0.100	0.410	492.1	2022.0
-7.3498816	57.27197	0.100	0.410	492.1	2022.0
-7.1242865	57.672283	0.137	0.153	674.9	753.6
-7.289001	57.163076	0.207	0.076	1023.4	372.5

FIGURE LEGENDS

Figure 1. (a) Prior expectation and actual value of σ (in km), and (b) fitted sigma versus the actual sigma. (c) shows the error in σ for one to five sources; positive errors represent fitted values greater than the true value, and negative errors fitted values lower than the true value. In both (a) and (b), replicates from single sources are shown in red, and from multiple (2-5) sources in black, with fitted lines in red. The red line in (b) shows the 1:1 line and the blue line the fitted line.

Figure 2. Geoprofiles showing the results of the GP analysis. Black dots show the locations of successful traps in (a) and the locations of mink sightings in (b). Contours show the top 50% of the geoprofile in both cases, in bands of 2%, with lighter colours corresponding to higher parts of the geoprofile. Sources (known den locations) are shown by the red squares.

FIGURES

FIGURE 1

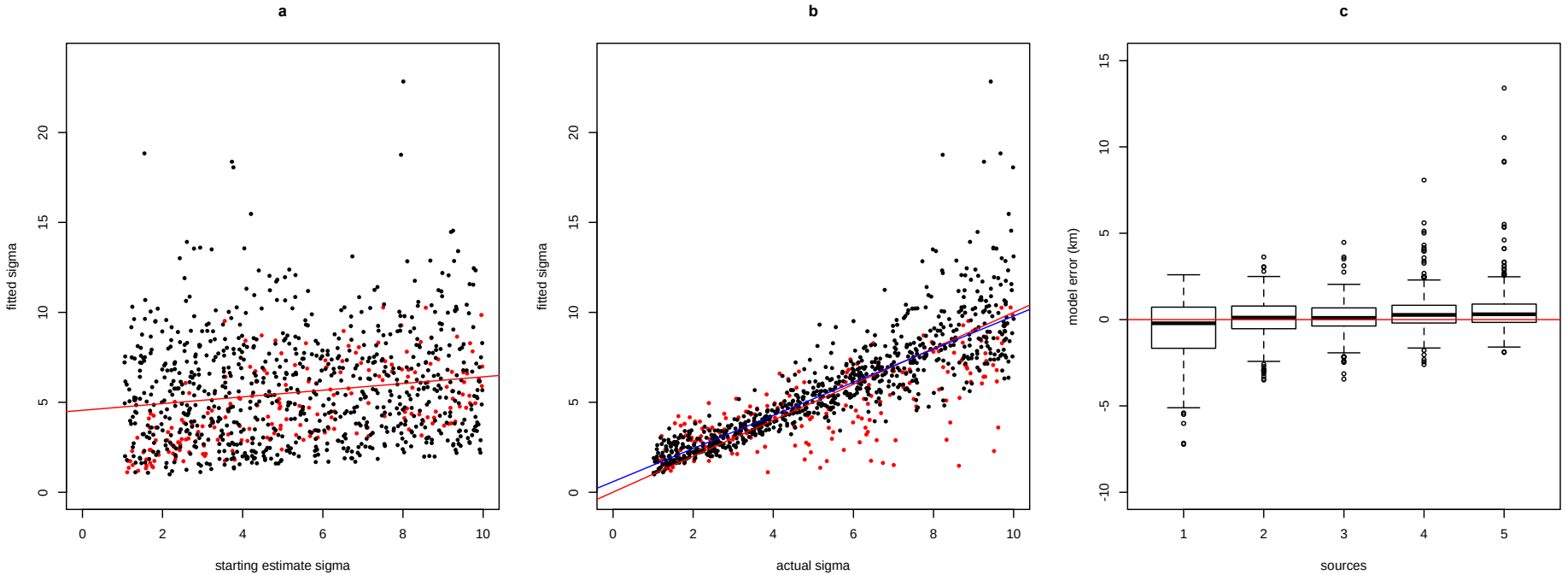
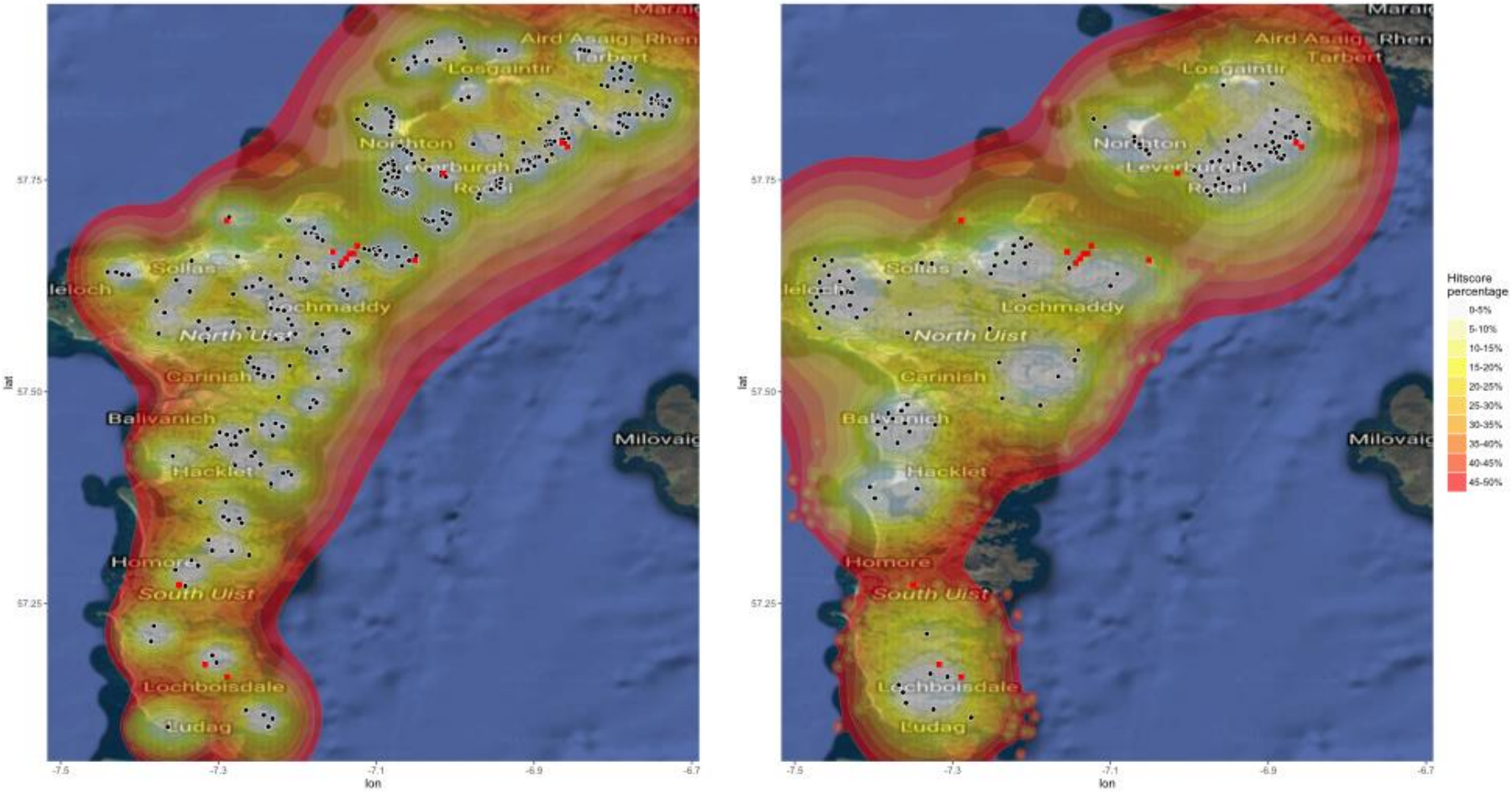


FIGURE 2


Losgaintir
Northton
Leverburgh
Rodel
Solas
Lochmaddy
North Uist
Carinish
Ballivanich
Hacklet
Milovaig
Homore
South Uist
Lochboisdale
Ludag
lat
lon
57.75
57.50
57.25
-7.5
-7.3
-7.1
-6.9
-6.7
Hitscore percentage
0-5%
5-10%
10-15%
15-20%
20-25%
25-30%
30-35%
35-40%
40-45%
45-50%